



## **MLflow**

### NetApp Solutions

NetApp  
August 14, 2024

# Sommario

- MLflow ..... 1
- Distribuzione MLflow ..... 1
- Tracciabilità tra set di dati e modelli con NetApp e MLflow ..... 2

# MLflow

## Distribuzione MLflow

Questa sezione descrive le attività che è necessario completare per implementare MLflow nel cluster Kubernetes.



È possibile implementare MLflow su piattaforme diverse da Kubernetes. L'implementazione di MLflow su piattaforme diverse da Kubernetes non rientra nell'ambito di questa soluzione.

### Prerequisiti

Prima di eseguire l'esercizio di implementazione descritto in questa sezione, si presuppone che siano già state eseguite le seguenti attività:

1. Hai già un cluster Kubernetes funzionante.
2. Hai già installato e configurato NetApp Astra Trident nel tuo cluster Kubernetes. Per ulteriori informazioni su Astra Trident, fare riferimento alla "[Documentazione di Astra Trident](#)".

### Installare Helm

MLflow è implementato usando Helm, un noto gestore di pacchetti per Kubernetes. Prima di implementare MLflow, è necessario installare Helm sul nodo di controllo Kubernetes. Per installare Helm, seguire la "[istruzioni per l'installazione](#)" nella documentazione ufficiale di Helm.

### Impostare la classe di storage Kubernetes predefinita

Prima di implementare MLflow, devi indicare una classe storage predefinita all'interno del cluster Kubernetes. Per designare una StorageClass predefinita all'interno del cluster, segui le istruzioni riportate nella "[Implementazione di Kubeflow](#)" sezione. Se è già stata designata una StorageClass predefinita all'interno del cluster, è possibile saltare questo passaggio.

### Distribuire MLflow

Una volta soddisfatti i prerequisiti, è possibile iniziare con la distribuzione di MLflow utilizzando il grafico del timone.

#### Configurare la distribuzione di MLflow Helm Chart.

Prima di distribuire MLflow utilizzando il grafico Helm, è possibile configurare la distribuzione in modo che utilizzi la classe di archiviazione NetApp Trident e modificare altri parametri in base alle proprie esigenze utilizzando un file **config.yaml**. Un esempio di file **config.yaml** è disponibile all'indirizzo: <https://github.com/bitnami/charts/blob/main/bitnami/mlflow/values.yaml>



È possibile impostare Trident storageClass nel parametro **global.defaultStorageClass** del file **config.yaml** (ad esempio storageClass: "ontap-flexvol").

#### Installazione della Helm Chart

Il grafico Helm può essere installato con il file personalizzato **config.yaml** per MLflow utilizzando il seguente

comando:

```
helm install oci://registry-1.docker.io/bitnamicharts/mlflow -f  
config.yaml --generate-name --namespace jupyterhub
```



Il comando implementa MLflow sul cluster Kubernetes nella configurazione personalizzata tramite il file **config.yaml** fornito. MLflow viene distribuito nello spazio dei nomi specificato e un nome di release casuale viene assegnato tramite kuPensioni per la release.

## Controllare distribuzione

Una volta completata la distribuzione del grafico Helm, è possibile verificare se il servizio è accessibile utilizzando:

```
kubectl get service -n jupyterhub
```



Sostituire **jupyterhub** con lo spazio dei nomi utilizzato durante la distribuzione.

Vengono visualizzati i seguenti servizi:

NAME	TYPE	CLUSTER-IP	EXTERNAL-IP
mlflow-1719843029-minio	ClusterIP	10.233.22.4	<none>
80/TCP, 9001/TCP			
mlflow-1719843029-postgresql	ClusterIP	10.233.5.141	<none>
5432/TCP			
mlflow-1719843029-postgresql-hl	ClusterIP	None	<none>
5432/TCP			
mlflow-1719843029-tracking	NodePort	10.233.2.158	<none>
30002:30002/TCP			
25d			



Il file config.yaml è stato modificato per utilizzare il servizio NodePort per accedere a MLflow sulla porta 30002.

## Accesso a MLflow

Una volta che tutti i servizi relativi a MLflow sono in funzione, è possibile accedervi utilizzando l'indirizzo IP di NodePort o LoadBalancer specificato (ad es. <http://10.61.181.109:30002>)

## Tracciabilità tra set di dati e modelli con NetApp e MLflow

```
https://github.com/NetApp/netapp-dataops-  
toolkit/tree/main/netapp_dataops_k8s["NetApp DataOps Toolkit per  
Kubernetes"^]Può essere utilizzato insieme alle funzionalità di  
tracciamento degli esperimenti di MLflow per implementare la tracciabilità  
da set di dati a modello o da spazio di lavoro a modello.
```

Per implementare la tracciabilità da set di dati a modello o da spazio di lavoro a modello, è sufficiente creare uno snapshot del set di dati o del volume di lavoro utilizzando DataOps Toolkit come parte dell'esecuzione del training, come illustrato nel seguente frammento di codice di esempio. Questo codice salverà il nome del volume di dati e il nome dello snapshot come tag associati alla sequenza di training specifica che si sta accedendo al server di tracciamento degli esperimenti MLflow.

```
...  
from netapp_dataops.k8s import create_volume_snapshot  
  
with mlflow.start_run() :  
    ...  
  
    namespace = "my_namespace" # Kubernetes namespace in which dataset  
    volume PVC resides  
    dataset_volume_name = "project1" # Name of PVC corresponding to  
    dataset volume  
    snapshot_name = "run1" # Name to assign to your new snapshot  
  
    # Create snapshot  
    create_volume_snapshot(  
        namespace=namespace,  
        pvc_name=dataset_volume_name,  
        snapshot_name=snapshot_name,  
        printOutput=True  
    )  
  
    # Log data volume name and snapshot name as "tags"  
    # associated with this training run in mlflow.  
    mlflow.set_tag("data_volume_name", dataset_volume_name)  
    mlflow.set_tag("snapshot_name", snapshot_name)  
  
    ...
```

## Informazioni sul copyright

Copyright © 2024 NetApp, Inc. Tutti i diritti riservati. Stampato negli Stati Uniti d'America. Nessuna porzione di questo documento soggetta a copyright può essere riprodotta in qualsiasi formato o mezzo (grafico, elettronico o meccanico, inclusi fotocopie, registrazione, nastri o storage in un sistema elettronico) senza previo consenso scritto da parte del detentore del copyright.

Il software derivato dal materiale sottoposto a copyright di NetApp è soggetto alla seguente licenza e dichiarazione di non responsabilità:

IL PRESENTE SOFTWARE VIENE FORNITO DA NETAPP "COSÌ COM'È" E SENZA QUALSIVOGLIA TIPO DI GARANZIA IMPLICITA O ESPRESSA FRA CUI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, GARANZIE IMPLICITE DI COMMERCIALIZZABILITÀ E IDONEITÀ PER UNO SCOPO SPECIFICO, CHE VENGONO DECLINATE DAL PRESENTE DOCUMENTO. NETAPP NON VERRÀ CONSIDERATA RESPONSABILE IN ALCUN CASO PER QUALSIVOGLIA DANNO DIRETTO, INDIRETTO, ACCIDENTALE, SPECIALE, ESEMPLARE E CONSEGUENZIALE (COMPRESI, A TITOLO ESEMPLIFICATIVO E NON ESAUSTIVO, PROCUREMENT O SOSTITUZIONE DI MERCI O SERVIZI, IMPOSSIBILITÀ DI UTILIZZO O PERDITA DI DATI O PROFITTI OPPURE INTERRUZIONE DELL'ATTIVITÀ AZIENDALE) CAUSATO IN QUALSIVOGLIA MODO O IN RELAZIONE A QUALUNQUE TEORIA DI RESPONSABILITÀ, SIA ESSA CONTRATTUALE, RIGOROSA O DOVUTA A INSOLVENZA (COMPRESA LA NEGLIGENZA O ALTRO) INSORTA IN QUALSIASI MODO ATTRAVERSO L'UTILIZZO DEL PRESENTE SOFTWARE ANCHE IN PRESENZA DI UN PREAVVISO CIRCA L'EVENTUALITÀ DI QUESTO TIPO DI DANNI.

NetApp si riserva il diritto di modificare in qualsiasi momento qualunque prodotto descritto nel presente documento senza fornire alcun preavviso. NetApp non si assume alcuna responsabilità circa l'utilizzo dei prodotti o materiali descritti nel presente documento, con l'eccezione di quanto concordato espressamente e per iscritto da NetApp. L'utilizzo o l'acquisto del presente prodotto non comporta il rilascio di una licenza nell'ambito di un qualche diritto di brevetto, marchio commerciale o altro diritto di proprietà intellettuale di NetApp.

Il prodotto descritto in questa guida può essere protetto da uno o più brevetti degli Stati Uniti, esteri o in attesa di approvazione.

LEGENDA PER I DIRITTI SOTTOPOSTI A LIMITAZIONE: l'utilizzo, la duplicazione o la divulgazione da parte degli enti governativi sono soggetti alle limitazioni indicate nel sottoparagrafo (b)(3) della clausola Rights in Technical Data and Computer Software del DFARS 252.227-7013 (FEB 2014) e FAR 52.227-19 (DIC 2007).

I dati contenuti nel presente documento riguardano un articolo commerciale (secondo la definizione data in FAR 2.101) e sono di proprietà di NetApp, Inc. Tutti i dati tecnici e il software NetApp forniti secondo i termini del presente Contratto sono articoli aventi natura commerciale, sviluppati con finanziamenti esclusivamente privati. Il governo statunitense ha una licenza irrevocabile limitata, non esclusiva, non trasferibile, non cedibile, mondiale, per l'utilizzo dei Dati esclusivamente in connessione con e a supporto di un contratto governativo statunitense in base al quale i Dati sono distribuiti. Con la sola esclusione di quanto indicato nel presente documento, i Dati non possono essere utilizzati, divulgati, riprodotti, modificati, visualizzati o mostrati senza la previa approvazione scritta di NetApp, Inc. I diritti di licenza del governo degli Stati Uniti per il Dipartimento della Difesa sono limitati ai diritti identificati nella clausola DFARS 252.227-7015(b) (FEB 2014).

## Informazioni sul marchio commerciale

NETAPP, il logo NETAPP e i marchi elencati alla pagina <http://www.netapp.com/TM> sono marchi di NetApp, Inc. Gli altri nomi di aziende e prodotti potrebbero essere marchi dei rispettivi proprietari.